Constraint Grammars

Thierry Martinez

Acknowledgments to Rémy Haemmerlé for the original impulse to this work.

Contraintes Project–Team, INRIA Paris–Rocquencourt, France

21 July 2008
Outline

1. Introduction.


4. Conclusions & Perspectives.
Constraint Sets.

Observational confluence

\[
\begin{align*}
\text{nil}(L) \setminus \text{positive_list}(L) & \iff \text{true}. \\
\text{list}(L, H, T) \setminus \text{positive_list}(L) & \iff H > 0, \text{positive_list}(T).
\end{align*}
\]

- Confluent for queries modelling a list (first arguments of list tokens are pair-wise distincts).
- Non-confluent for queries modelling a tree (with two list tokens sharing a common first argument).

Characterizing (L)CC Stores

Describing Global Constraints

Given a set of constraints \( S \):

- either there exists a satisfied constraint in \( S \):

\[
\text{REGULAR}(x_1, \ldots, x_n, L) \text{ with } S = \{x_1 = u_1 \land \cdots \land x_n = u_n\}_{u \in L}
\]

- or all the constraints of \( S \) are satisfied:

\[
\text{ALL-DIFFERENT}(x_1, \ldots, x_n) \text{ with } S = \{x_i \neq x_j\}_{i \neq j}
\]
Linear Constraint Systems.

Let $V$ be a set of variables.

- If $C$ is a set of formulae over the set of variables $V$, let $\mathcal{C}[C]$ be the closure of $C$ by conjunction ($\otimes$) and variable hiding ($\exists$).
- If $\vdash$ is a binary relation over $C$, let $\widehat{\vdash}$ be the closure of $\vdash$ by the rules of linear logic.

**Definition**

A (linear) **constraint system** is a pair $(C,\vdash)$ where:

- $C$ is a set of formulae closed by conjunction ($\otimes$) and variable hiding ($\exists$);
- $\vdash \subseteq C^2$ is a relation closed by the rules of linear logic.
Examples of Constraint Systems

Let $\Sigma$ be a signature. Let $t(\Sigma) \doteq \{ f(v_1, \ldots, v_n) | f / n \in \Sigma \}$.

**linear-token system**

The **linear-token system** over $\Sigma$ is the constraint system $\text{TOK}(\Sigma) \doteq (C[t(\Sigma)], \emptyset)$.

**linear-token system with equality**

The **linear-token system with equality** over $\Sigma$ is the constraint system $\text{TOK}_=(\Sigma) \doteq (C[t(\Sigma)] \cup \{ !(x = y) | x, y \in V \}, \models)$, where $\models$ is the smallest relation with the axioms of the equality theory:

- $\models!(x = x)$;
- $!(x = y) \models!(y = x)$;
- $!(x = y) \otimes !(y = z) \models !(x = z)$;
- $f(x_0, \ldots, x_n) \otimes !(x_0 = y_0) \otimes \cdots \otimes !(x_n = y_n) \models f(y_0, \ldots, y_n)$
Context-Free Constraint Grammars (CFCG).

**Definition**

A context-free constraint grammar (CFCG) is a tuple \((V, C, \vdash, \Sigma, P)\) where:

- \(V\) is a set of variables and \((C, \vdash)\) is a constraint-system over \(V\);
- \(\Sigma\) is the signature of non-terminal symbols:
  \[N \doteq \{ f(y_1, \ldots, y_n) | f/n \in \Sigma \text{ and } y_1, \ldots, y_n \in V \text{ pair-wise distincts}\}\]
- \(P \subseteq N \times \overline{C}\) is the set of productions, with \(\overline{C} \doteq \overline{C}[C \cup N]\).
  Every production \((f(y_1, \ldots, y_n), u) \in P\) is such that \(\text{fv}(u) \subseteq \{y_1, \ldots, y_n\}\).
  A production is denoted: \(f(y_1, \ldots, y_n) ::= u\).

**Example:**

\[
\begin{align*}
h(X, Y) & ::= \text{edge}(X, Y) \\
h(X, Y) & ::= \exists Z. \text{edge}(X, Z) \otimes h(Z, Y)
\end{align*}
\]
Derivations for CFCGs.

Let \((V, C, \vdash, \Sigma, P)\) be a context-free constraint grammar. Let \(\overline{C} = C \cup N\).

**Definition**

Let \(\rightarrow \subseteq C^2\) be the smallest relation satisfying the following rules:

\[
\begin{align*}
  f(y_1, \ldots, y_n) &:= u \\
  (f(y_1, \ldots, y_n))\sigma &\rightarrow (u)\sigma \\
  u &\rightarrow v \\
  \exists x. u &\rightarrow \exists x v \\
  u \otimes w &\rightarrow v \otimes w \\
  w \otimes u &\rightarrow w \otimes v
\end{align*}
\]

A derivation is a sequence \((c_i)_i\) with elements in \(\overline{C}\) such that \(c_0 \rightarrow c_1 \rightarrow \ldots\).

**Example:**

\[
\begin{align*}
h(X, Y) &:= \text{edge}(X, Y) \\
h(X, Y) &:= \exists Z. \text{edge}(X, Z) \otimes h(Z, Y) \\
h(A, B) &\rightarrow \exists X_1. \text{edge}(A, X_1) \otimes h(X_1, B) \\
&\rightarrow \exists X_1 X_2. \text{edge}(A, X_1) \otimes \text{edge}(X_1, X_2) \otimes h(X_2, B) \\
&\rightarrow \exists X_1 X_2. \text{edge}(A, X_1) \otimes \text{edge}(X_1, X_2) \otimes \text{edge}(X_2, B)
\end{align*}
\]
Languages accepted by CFCGs.

Let $G \vdash (V, C, \vdash, \Sigma, P)$ be a context-free constraint grammar. Let $\overline{C} \vdash C[C \cup N]$.

**Definition**

The **constraint language** accepted by a schema $c_0 \in \overline{C}$ and the grammar $G$ is:

$$L_G(c_0) = \{ c \in C | (\exists n) (\exists (c_i)_{0 \leq i \leq n}) (c_0 \to \cdots \to c_n \land c_n \vdash \vdash c) \}$$

**Example:**

$$h(X, Y) ::= \text{edge}(X, Y)$$

$$h(X, Y) ::= \exists Z. \text{edge}(X, Z) \otimes h(Z, Y)$$

$$L_G(h(A, B)) = \{ \text{hamiltonian paths from } A \text{ to } B \}$$
Link with Hyperedge Replacement Graph Grammars.

Claim
Over linear token systems, context-free constraint grammars have the same expressive power than hyperedge replacement graph grammars.


Theorem (K.-J. Lange and E. Welzl 87)
*Hyperedge replacement graph language membership problem is NP-complete.*
Hypergraphs.

A **labelled set** $\hat{E}$ is a tuple $(E, E^\Sigma, E^\lambda)$, with $E^\Sigma$ alphabet and $E^\lambda : E \rightarrow E^\Sigma$.

**Definition**

A **hypergraph** is a tuple $(V, \hat{E}, S, \rightarrow)$, where:

- $V$ is the set of vertices; $\hat{E}$ is the labelled set of hyperedges;
- $S$ is the alphabet of selectors;
- $\rightarrow \subseteq E \times S \times V$ is the incidence relation. $(e, s, v) \in \rightarrow$ is denoted $e \overset{s}{\rightarrow} v$.
- $\rightarrow$ is functional: $(\forall e \ s \ v_1 \ v_2) e \overset{s}{\rightarrow} v_1 \land e \overset{s}{\rightarrow} v_2 \Rightarrow v_1 = v_2$

Vertices are represented as fat dots, hyperedges as boxes.

Let $\mathcal{H}(V, E, S)$ the set of hypergraphs over $V, E, S$. 

Let $\mathcal{H}(V, E, S)$ the set of hypergraphs over $V, E, S$. 

![Diagram of hypergraphs](image)
Embedding of Hypergraphs into Constraints.

Let $H = (V, \hat{E}, S, \rightarrow)$ be a hypergraph.

- For all $e \in E$, let $\text{dom}(e) = \{ s \in S | (\exists v \in V) (e \overset{s}{\rightarrow} v) \}$.
- Let $n(e) = \lvert \text{dom}(e) \rvert$ and $r_e : \{ 1, \ldots, n(e) \} \rightarrow \text{dom}(e)$.
- Let $\Sigma = \{ f / n \in E^2 \times \mathbb{N} | (\exists e \in E) (f = E^\lambda(e) \text{ and } n = n(e)) \}$.

**Definition**

The **constraint-embedding** of the hypergraph $h$ is the following constraint of $\text{TOK}(\Sigma)$:

$$[h] \doteq (\exists v_1 \ldots v_n) \left( \bigotimes_{e \in E} [E^\lambda(e)](y_1, \ldots, y_n) \right)$$

$s_1, s_2 : E \rightarrow S$

**Proposition**

For every closed constraint $c \in \text{TOK}(\Sigma)$, there exists a hypergraph $h$ such that

$$[h] \not\vdash c$$
Hypergraph Isomorphism.

Definition

Two hypergraphs \( h_0 \doteq (V, (E, E^\Sigma, E^\lambda_0), S, \rightarrow_0) \) and \( h_1 \doteq (V, (E, E^\Sigma, E^\lambda_1), S, \rightarrow_1) \) are \textbf{isomorphic}, denoted \( h_0 \simeq h_1 \), when there exist \( \sigma_V : V \leftarrow V \) and \( \sigma_E : E \leftarrow E \) such that:

1. \( (\forall e \in E)(E^\lambda_0(e) = E^\lambda_1(\sigma_E(e))) \);
2. \( (\forall e \in E)(\forall s \in S)(\forall v \in V)(e \rightarrow_0^s v \iff (\sigma_E(e) \rightarrow_1^s \sigma_V(v)) \).

Proposition

\[ h_0 \simeq h_1 \iff \llbracket h_0 \rrbracket \vdash \llbracket h_1 \rrbracket \]
Hyperedge Replacement Graph Grammars (HRGs).

Let $N, V, E, E^\Sigma, S$ be defined accordingly to the following definition. We denote $\mathcal{H}^+(N, V, E, E^\Sigma, S)$ the set of augmented hypergraphs, that is hypergraphs whose edges are labelled with $\hat{E}^\Sigma$ and vertices are in $\hat{V}$.

**Definition**

A **h-edge replacement graph grammar** is a tuple $(N, V, E, E^\Sigma, S, P)$, with:

- $N$ is the alphabet of non-terminals, such that $E^\Sigma \cap N = \emptyset$; $\hat{E}^\Sigma \equiv E^\Sigma \cup N$;
- $V$ is the set of vertices;
- $E$ is the set of hyperedges and $E^\Sigma$ the set of ground labels;
- $S$ is the alphabet of selectors, such that $V \cap S = \emptyset$; $\hat{V} \equiv V \cup S$;
- $P \subseteq N \times \wp(S) \times \mathcal{H}^+(N, V, E, E^\Sigma, S)$ is the set of productions.
Derivations for HRGs.

Let \((N, V, E, E^\Sigma, S, P)\) be a hyperedge replacement graph grammar.

**Definition**

Let \(\rightarrow \subseteq \mathcal{H}^+(N, V, E, E^\Sigma, S)^2\) be the relation such that:

\((V_0, E_0, S, \rightarrow_0) \rightarrow (V_1, E_1, S, \rightarrow_1)\) iff there exist:

- a hyperedge \(e \in E_0\); a production \((n, S_p, (V_p, E_p, S, \rightarrow_p)) \in P\);
- \(\sigma^E_{0\rightarrow 1} : E_0 \setminus \{e\} \rightarrow E_1, \sigma^V_{0\rightarrow 1} : V_0 \rightarrow V_1, \sigma^E_{p\rightarrow 1} : E_p \rightarrow E_1, \sigma^V_{p\rightarrow 1} : V_p \setminus S \rightarrow V_1\) such that:
  - \(E_0^\lambda(e) = n; \text{im}(\sigma^E_{0\rightarrow 1}) \uplus \text{im}(\sigma^E_{p\rightarrow 1}) = E_1\) and \(\text{im}(\sigma^V_{0\rightarrow 1}) \uplus \text{im}(\sigma^V_{p\rightarrow 1}) = V_1\);
  - \((\forall e_0 \in E_0 \setminus \{e\})(\forall e_p \in E_p)(\forall s s' \in S)(\forall v_0 \in V_0)(\forall v_p \in V_p \setminus S)\)

\[
E_0^\lambda(e_0) = E_1^\lambda(\sigma^E_{0\rightarrow 1}(e_0))
\]

\[
E_p^\lambda(e_p) = E_1^\lambda(\sigma^E_{p\rightarrow 1}(e_p))
\]

\[
e_0 \xrightarrow{s_0} v_0 \iff \sigma^E_{0\rightarrow 1}(e_0) \rightarrow^{s_1} \sigma^V_{0\rightarrow 1}(v_0)
\]

\[
e_p \xrightarrow{s_p} v_p \iff \sigma^E_{p\rightarrow 1}(e_p) \rightarrow^{s_1} \sigma^V_{p\rightarrow 1}(v_p)
\]

\[
e \xrightarrow{s_0} v_0 \wedge e_p \xrightarrow{s'_p} s \iff \sigma^E_{p\rightarrow 1}(e_p) \rightarrow^{s'_1} \sigma^V_{0\rightarrow 1}(v_0)
\]
Embedding of HRGs into CFCGs.

Let $G = (N, V, E, E^\Sigma, S, P)$ be a hyperedge replacement graph grammar.

1. $[\bullet]$ can be extended to an embedding of augmented hypergraphs into scheme of $\overline{C}$, with $N = \{f(y_1, \ldots y_n) | (\exists h)((f, \{y_1, \ldots, y_n\}, h) \in P)\}$;

2. therefore $[\bullet]$ can be extended to an embedding of HRG productions into context-free constraint productions;

3. therefore $[\bullet]$ can be extended to an embedding of HRG grammars into context-free constraint grammars.

Definition

The **hypergraph language** accepted by an augmented hypergraph $h_0 \in \mathcal{H}^+(N, V, E, E^\Sigma, S)$ and the grammar $G$ is:

$$L_G(h_0) = \{h \in \mathcal{H}(V, E, S) | (\exists n)(\exists (h_i)_{0 \leq i \leq n})(h_0 \rightarrow \cdots \rightarrow h_n \land h \simeq h_n)$$

Proposition

*For all $h_0 \in \mathcal{H}^+(N, V, E, E^\Sigma, S)$, we have $[L_G(h)] = L_{[G]}([h])$.***
With Equality Theory.

**Lemma:** for all constraint \( c \in \text{TOK}_= (\Sigma) \), there exists a constraint \( c' \in \text{TOK}(\Sigma) \subseteq \text{TOK}_= (\Sigma) \) such that \( c \models c' \).

**Proposition**

*For every closed constraint \( c \in \text{TOK}_= (\Sigma) \), there exists a hypergraph \( h \) such that*

\[
[h] \models c
\]

**Lemma:** For a schema \( c \in \overline{\text{TOK}_= (\Sigma)} \) and a grammar \( G \) over \( \text{TOK}_= (\Sigma) \), there exists \( c' \in \overline{\text{TOK}(\Sigma)} \) and a \( G' \) over \( \text{TOK}(\Sigma) \) such that:

\[
L_G(c) = \{ h \in \text{TOK}_= (\Sigma) | \exists h' \in L_{G'}(c'), h \models h' \}
\]

**Proposition**

*Over \( \text{TOK}_= (\Sigma) \), context-free constraint language membership problem is NP-complete.*
Applications to Global Constraints.

Let $A$ be an alphabet. Every $u = u_1 \cdots u_n \in A^*$ can be represented as a string-hypergraph:

\[
\begin{array}{cccccc}
 & 0 & u_1 & 1 & \cdots & 0 & u_n & 1 & \\
\end{array}
\]

**Proposition (Habel and Kreowski 87)**

*Every context-free string language $G$ can be recognized by a hyperedge replacement graph grammar $[G]$.***

**Corollary**: Every context-free string language $G$ can be recognized by a context-free constraint grammar $[G]$.

**Proposition**

$\text{CONTEXT-FREE}(x_1, \ldots, x_n, G) \equiv (\exists) [G] (x_1, \ldots, x_n)$
Limitations.

**Definition**

A **jungle** is a hypergraph \((V, \dot{E}, \{\text{in}_k, \text{out}\}, \rightarrow)\) such that:

- \((\forall v \in V) \left( \left| \{ e \in E | e^{\text{out}} \rightarrow v \} \right| \leq 1 \right)\);
- \(\rightarrow\) is acyclic.


“**Corollary 2:** CHR folding is terminating and confluent.”

*Only true if the query is a jungle!*

**Proposition**

\[ \{ h \in \mathcal{H}(V, E, S) | h \text{ is a jungle} \} \text{ is not a hyperedge replacement graph language.} \]
Domain-Sensitive Constraint Grammars. (1/2)

A **two-sorted signature** $\Sigma$ is a subset of $F \times \mathbb{N} \times \mathbb{N}$, where $F$ is a countable symbol set. Items $(f, n, m) \in \Sigma$ are denoted $f / (n, m)$.

Let $V$ and $W$ be two set of variables and $L$ a set of labels. The **set constraint system** is the smallest classical constraint system $(S, \vdash_S)$ such that:

\[
\begin{align*}
\frac{w \in W}{w = \emptyset \in S} & \quad \frac{v \in V \quad w \in W}{w = \{v\} \in S} & \quad \frac{v \in V \quad w \in W}{(v \in w) \in S} & \quad \frac{w \in W \quad l \in L}{w = f v l \in S} \\
\frac{w_1 \in W \quad w_2 \in W \quad w_3 \in W \quad \text{op} \in \{\cap, \cup, \setminus\}}{w_1 = w_2 \ \text{op} \ w_3 \in S} & \quad & & \\
\end{align*}
\]

with $w = \{v\} \vdash_S v \in w \quad w = \emptyset \vdash_S v \notin w$

and $v \in w_1 \ \text{op}_1 \ v \in w_2 \vdash_S v \in w_1 \ \text{op}_2 \ w_2$ with

$(\text{op}_1, \text{op}_2) \in \{((\land, \cap)); ((\lor, \cup)); ((\land \neg, \cap))\}$...
Definition

A **domain-sensitive constraint grammar** (DSCG) is a tuple 
\((V, W, L, C, \vdash, \Sigma, P)\) where:

- \(V, W\) are sets of variables, \(L\) a set of labels;
- \((C, \vdash)\) is a constraint-system over \(V\);
- \(\Sigma\) is the two-sorted signature of non-terminal symbols:
  \[ N \doteq \left\{ l : f(y_1, \ldots, y_n, Y_1, \ldots, Y_m) \mid l \in L, f / (n, m) \in \Sigma \text{ and } y_1, \ldots, y_n \in V, Y_1, \ldots, Y_m \in W \text{ p.-w. dist.} \right\} \]
- \(P \subseteq N \times \overline{C} \times S\) is the set of productions, with \(\overline{C} \doteq \mathbb{C}[C \cup N]\).

A production is denoted: 
\[ f(y_1, \ldots, y_n, Y_1, \ldots, Y_m) \ ::= u \mid g. \]
Applications of DSCGs.

**Jungles** Let $\Sigma$ be a signature. For all $f/n \in \Sigma$:

$$j(r) ::= \text{eq}(r, f(x_1, \ldots, x_n)) \otimes l_1 : j(x_1) \otimes \cdots \otimes l_n : j(x_n) | r \not\in \text{fv}(l_1) \cup \cdots \cup \text{fv}(l_n)$$

**all-different**

$$a(S) ::= x_1 \neq x_2 | x_1 \in S \land x_2 \in S \setminus \{x_1\}$$

$$\text{ALL-DIFFERENT}(x_1, \ldots, x_n) \equiv (\forall) a(\{x_1, \ldots, x_n\})$$
Conclusions & Perspectives.

Objective

- Describing set of constraints;
- for observational confluence;
- for defining global constraints.

Contributions

- Definition of constraint grammar;
- as expressive as hyperedge-replacement graph grammar;
- much more easy to define;
- more general.

Future work

- Decidability and complexity analysis for other constraint systems.